AN EXAMINATION OF INTROGRESSION IN THE *TRILLIUM ERECTUM* SPECIES COMPLEX USING MICROSATELLITE ANALYSIS

A thesis presented to the faculty of the Graduate School of Western Carolina University in partial fulfillment of the requirements for the degree of Master of Science in Biology

By

Austin William Brenek

Director: Dr. Katherine Gould Mathews
Professor of Biology
Biology Department

Committee Members: Dr. Beverly Collins, Biology,
Dr. Luiz Lima Da Silveira, Biology

April 2021

ACKNOWLEDGEMENTS

I would like to thank Mr. Paul Super for aiding in my access to my field sites in the Great Smoky Mountains National Park.
I would like to thank my committee for their insight on how to address issues that occurred during my study.
My advisor Dr. Kathy Mathews who has been with me at every step of my project.
Dark Cove Farm for allowing me access to their property which is absolutely covered in *Trillium*.
And my buddy David who has been a source of inspiration since we met.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

AN EXAMINATION OF INTROGRESSION IN THE *TRILLIUM ERECTUM* SPECIES COMPLEX USING MICROSATELLITE ANALYSIS

Austin Brenek, M.S. Biology

Western Carolina University (April 2021)

Director: Dr. Kathy Mathews

There are seven named taxa in the *Trillium erectum* L. species complex native to North America, many of which are experiencing secondary contact and hybridizing due to a lack of reproductive isolating mechanisms. This project will focus on *T. erectum* var. *album*, a white-flowered taxon, *T. erectum* var. *erectum*, a red-flowered taxon, and *T. rugelii*, another white-flowered taxon, each of which occurs in the southern Appalachian Mountains in populations that overlap in both geographic distribution and flowering phenology. Using three microsatellite loci developed for a related *Trillium* sp., this study examines the hybridization and genetic structure of several populations of the *T. erectum* species complex located in three counties in the western region of North Carolina to determine if the taxa growing in mixed populations are 1) interbreeding and 2) if so, quantify the amount of admixture in each population to examine what factors (taxon identity, geographic range, or flower color) are most influencing hybridization. Allele frequency analyses of microsatellite loci were also used to compare the populations, and Principle Components Analysis was used to make pairwise comparisons of both the fixation index and Nei's Genetic Distance calculations for each geographic population. Finally, structure analysis was used to identify populations and quantify admixture based on allele frequency and assumptions of gene flow using Bayesian statistical methods. Results from population genetics

and genetic structure analyses suggest that allele sharing occurs primarily based on taxon identity and geographic proximity, but flower color may also play a role in influencing gene flow.

CHAPTER ONE: INTRODUCTION

Background

Questions in the field of evolutionary biology often focus on speciation and the processes by which new species originate. Hybridization is one such process; hybrids are formed when two unique species mate and produce offspring. If the offspring are fertile and the hybrid offspring are isolated, this process can result in speciation. This study examines hybridization in the *Trillium erectum* L. species complex, a group of pedicellate *Trillium* that are known to hybridize, to find out if different taxa growing together in mixed populations are interbreeding, and if so, quantify the amount of admixture and examine what factors are influencing hybridization. I accomplish this by examining gene flow, performing genetic analysis on several mixed taxa and single taxon populations of species in the *Trillium erectum* complex that are known to hybridize. Research questions I address are:

1) What can allelic data from microsatellite markers tell us about the extent of hybridization in some members of the *T. erectum* complex?
2) What factors are most correlated with hybridization: geographic proximity, species identity or pollinator selection based on flower color?

Organism Description

There are seven named taxa of the *Trillium erectum* L. species complex that are native to North America (Case & Case 1997). This study will focus on *T. erectum* var. *album*, *T. erectum* var. *erectum*, and *T. rugelii*, three taxa that occur in the southern Appalachian Mountains and are found in populations that overlap in geographic range and flowering phenology (Case and Case

1997, Stoehrel 2010). The taxa are part of the species complex as they share many characteristics and boundaries between them are unclear (Case and Case 1997). Within the complex, flower color varies from shades of red to white, and taxa may have two opposite color morphs that can be found in allopatry or sympatry, as in the two varieties of *T. erectum*: *T. erectum* var. *album* has white flowers and *T. erectum* var. *erectum* has red flowers. *Trillium rugelli* typically has white flowers but presumed hybrid forms have red or pink flowers. Seeds are dispersed primarily by ants but have also been observed to be dispersed by deer (Vellend et al 2003).

Inbreeding is also prevalent as studies have shown that *T. erectum* lacks a self-incompatibility mechanism (Sage et al. 2001). Furthermore, *Trillium* also reproduce clonally by producing multiple flowers from a single rhizome (Case and Case 1997). Both of these aspects limit genetic diversity as they are alternatives to sexual reproduction that produce offspring that are identical to the parent.

Hybridization

Hybridization is known to occur among some recently evolved plant species because of the lack of genetic isolating mechanisms (e.g., Furches et al. 2013, Khosravi et al. 2013, Yan et al. 2017). A general mechanism that mitigates the effect of hybridization on populations is the tendency of organisms to produce unfertile offspring (e.g., Todesco et al. 2016, Yan et al. 2017). The introgression of genes and formation of viable hybrid offspring in closely related plants is common, however, it may also increase genetic diversity when genes are shared between similar sized populations (Furches et al. 2013).

However, in some cases, hybridization can lead to a loss of genetic diversity, leading to the potential loss of rare species (Yan et al. 2017). This process is known as genetic swamping

and occurs when genes are being transferred between large and small populations of individuals that produce viable hybrid offspring. Over time, the process of fixation, the convergence of a population's genotypic diversity, can lead to the loss of the genetic identity of the smaller population (Todesco et al. 2016, Zaya et al. 2015). Although the rare species may still be producing offspring, over time, the genetic identity of the species is diluted to the point of extinction. In this scenario the rare species may be less fit for its environment or is occupying a niche that the more abundant species is also being forced to migrate into. This process is often exacerbated by climate change as species are often forced to shift geographic location potentially causing similar taxa that lack genetic isolating mechanisms to experience secondary contact (Todesco et al. 2016), which may be occurring in the *T. erectum* complex.

Using a molecular clock model, Millam (2006) suggested the *T. erectum* complex formed no more than 900,000 years ago during the Middle Pleistocene, with a division of the complex occurring after glaciers receded. However, the study does not conclude that the division of taxa is a result of speciation or hybridization among members of the complex (Millam 2006). Due to a lack of reproductive isolating mechanisms, species of *Trillium* in the Sothern Appalachian Mountains that may be experiencing secondary contact with closely related taxa can hybridize and produce fertile offspring (Stoehrel 2010). Models predict the loss of genetic variation in recently colonized derived populations encountering secondary contact with reproducibly compatible species (Ibrahim et al. 1995). To better understand the viability and genetic diversity of closely related taxa in the *T. erectum* complex, it is important to understand the population structure and evolutionary processes that are affecting it. These processes typically include natural selection, genetic drift, and gene flow (Griffin and Barrett 2004). My study examines the

amount of gene flow occurring among nine populations of two named varieties of one species, *T. erectum* var. *erectum* and *T. erectum* var. *album*, and a second species, *T. rugellii*.

The reason I am studying these three taxa is that in the Southern Appalachian region mixed taxon populations are a common occurrence, and in some of these populations both red and white flowered individuals of the same species exist in close proximity, which leads to the question of what function different colored flowers serves (see Pollination section below). Furthermore, putative hybrids between both varieties of *T. erectum* and *T. rugellii* have been observed. Historically, the *T. erectum* species complex has been under taxonomic scrutiny and as of yet relationships among taxa have yet to be resolved. Knowing the mechanisms that allow mixed populations of both distinct varieties as well as hybridizing individuals to exist can provide insight on taxonomic classification as well as the evolutionary context in which these taxa have arisen.

Species and Taxonomic Concepts

When considering the classification of taxa in the *T. erectum* species complex, it necessary to consider the concept of species especially when considering the complex nature of the relationship within the complex. The Biological Species Concept distinguishes species by the capability of individuals to interbreed and produce fertile offspring (DeQueiroz 2007). The Evolutionary Species Concept classifies species by evolutionary lineage, or ancestry, regardless of morphological characteristics (Wiley 1978). The Morphological Species Concept differentiates species by differences in morphological characteristics (Wiley 1978). These concepts may be applied to the taxa of *Trillium* studied here by considering the genetic data gathered and the analyses of gene flow conducted.

Taxonomic concepts and their relationships are important for understanding the historical aspects of nomenclature, how taxa have been previously classified, and which taxonomic entities (such as previously described species, varieties, or subspecies) are encompassed in a name when used in a flora or revision. Taxonomic concepts can resolve semantic arguments when the placement of taxa have undergone multiple revisions, and allow us to describe and classify organisms while avoiding disagreements between past classifications (Franz and Peet 2009). In this study, when I refer to the taxa of *T. erectum* var. *erectum*, *T. erectum* var. *album*, and *T. rugelii*, I refer to them in the concepts described by Weakley (2020).

Historically there has been much discrepancy in the classification and position of the taxa within the *T. erectum* species complex (Barksdale 1939, Osaloo et al. 1999, Milliam 2006). Previous taxonomic concepts of the two named varieties of *T. erectum* only differ in the acceptance of the two color morphologies as varieties or simply a difference in phenotype (Weakley 2020). The concept for both *T. erectum* var. *erectum* and *T. erectum* var. *album* is congruent with the concept of *T. erectum* described in the Flora of North America Vol. 26 (Case in FNA26 1993) as each of these describes *T. erectum* var. *album* as a distinct variety (Weakley 2020). These concepts are less inclusive than Case and Case (1997), Patrick (1986), Patrick (2007), which are themselves less inclusive than the concept of *T. erectum* Linnaeus var. *erectum* described in Radford, Ahles, and Bell's Manual of the Vascular flora of the Carolinas (1968; Weakley 2020). Radford et al. (1968) recognize *T. erectum* but makes no distinction between the red and white varieties (Weakley 2020).

The *T. rugelii* concepts described in Weakley (2020) are congruent with the Flora of North America Vol. 26 (Case in FNA26 1993), Case and Case (1997), Patrick (1986), and Patrick (2007) and less inclusive than Radford, Ahles, and Bell's Manual of the Vascular flora of

the Carolinas (Radford et al. 1968). Which are less inclusive than the *Trillium cernuum* var. *marcranthum* concept described by A.J. Eames and Wiegand whose broad description included what is now referred to as *T. rugelii* (Weakley 2020).

Molecular Background

Molecular phylogenetic studies confirm that the taxa of the *T. erectum* species complex constitute a monophyletic group, although the relationships within the group are mostly unresolved (Kato et al. 1995, Osaloo 1999, Millam 2006). These studies also report that *T. camchatesens* is sister to the *T. erectum* complex (Kato et al. 1995). Other molecular studies of the *Trillium* genus have focused on sequencing plastid and nuclear genes that work well for inferring phylogeny in many groups of plant species. However, genetic relationships remain poorly resolved among members of the *T. erectum* complex due to the complex's high genetic similarity and recent divergence of the taxa (Osaloo 1999, Millam 2006, Stoehrel 2010).

Griffin and Barrett (2004) used higher resolution molecular markers to examine the genetic diversity of *T. erectum*. They examined polymorphisms at allozyme loci to find patterns in genetic variation and inbreeding among populations as well as to determine the relative rates of genetic diversity between southern and northern populations as the species is believed to have migrated north after the recession of glaciers (Griffin and Barrett 2004). The results of this study reported much higher levels of heterozygosity than expected given that populations are small and inbreeding should be prevalent (Griffin and Barrett 2004). They also found little evidence that would suggest a trend of greater genetic diversity from south to north, which might be expected by migration models (Griffin and Barrett 2004). This could be a result of the large number of compatible taxa that are found in overlapping ranges in the Southern Appalachians.

Millam conducted a review of informative molecular markers and suggested that simple sequence repeat markers may prove to be more informative in resolving low level phylogenetic relationships (2006). My study aims to add to the previous studies by using less conserved molecular markers such as simple sequence repeat microsatellite regions.

Microsatellites

Microsatellites are highly variable length differences among repeating nucleotide regions of DNA sequences. According to previous studies, these data have shown interspecies genetic diversity as well as determining rates of introgression between hybridizing species (Dhyani et al. 2020, Furches et al. 2013, Khosravi et al. 2013). Using rapidly mutating gene regions as markers, genetic variation can be analyzed among members of the same species as well as members of mixed species populations.

Microsatellites are repeat sequences of non-coding DNA that are formed by slippage, mutation, or unequal crossing over events that when repaired extend the length of the repeating regions (Page and Holmes 1998). The length of microsatellite regions is highly variable between organisms and when analyzed show differences between individuals of the same species. By use of these highly variable microsatellite markers, genetic "fingerprints" were generated for individuals of each taxon at each population site. With these data, I evaluated the genetic similarity among populations at different geographic locations and among different taxa by comparing percentage of heterozygosity of genotypes in populations, fixation index (Fst) and estimate the amount of genetic admixture between taxa using population structure analysis.

In order to isolate microsatellite markers from DNA sequences, primer pairs that flank the repeating region must be known (Page and Holmes 1998). Microsatellite primers were

developed for the sister species of the *T. erectum* complex, *T. camschatcense*, a species of *Trillium* occurring in the Honshu region of Japan (Kubota et al. 200). These markers have been tested in preliminary studies from WCU with positive results in isolating microsatellite regions of the *T. erectum* species complex (McCormick et al. 2018) Based on three microsatellite markers, hybridization appears to be extensive in some mixed-taxon populations (unpubl, K. Mathews), but additional populations were sampled to confirm these preliminary data.

Pollination

In the case of Angiosperms, the role of reproduction is often facilitated by an animal pollinator. The method in which a pollinator selects flowers influences gene flow in plant populations (Wu et al. 2018). This is because pollinators may seek out plants that produce flowers with a specific floral color, floral shape, or fragrance in search of rewards such as nectar. This process is known as pollinator mediated selection (Wu et al. 2018). The plant is attracting the pollinator and the preference of the pollinator influences gene flow through plant populations.

The amount of gene flow within or among species of mixed-species populations depends on whether random or non-random pollination is occurring, especially when populations are small (Page and Holmes 1998). When pollinators are selective of distinct characteristics of plants in mixed species sites, a mechanism known as pollinator selectivity sympatric speciation occurs as pollinators visit plants with certain characteristics, bringing pollen or genetic material along with them (Wu et al. 2018). In contrast, when pollination is random, members of populations may experience the introgression of genetic materials from other closely related species into the gene pool resulting in hybridization (Wu et al. 2018, Page and Holmes 1998).

In the case of the *T. erectum* species complex, flowers are insect pollinated; some species' flowers smell pleasant and sweet such as the *T. rugelii,* while other species, such as *T. erectum* have flowers that smell more like mushrooms or rancid meat that attract beetles and flies (Case and Case 1997). The varieties of *T. erectum* are known to be pollinated by beetles and flies; in a study of *T. erectum* pollinators in Vermont, Irwin found that the primary visitors that were observed to be carriers of *T. erectum* pollen were dipterans of the families Anthomyiidae, Sciaridae, and Sarcophagide, while less frequent pollinators were coleopterans (2000).

T. erectum var. *erectum* and *T. erectum* var. *album* are taxa that are found in sympatry in Jackson County, NC yet still maintain unique color morphology. This suggests that the transfer of alleles between individuals in theses populations is being segregated between the different color morphologies. Although the odor of the flower is thought to be the main trait responsible for attracting pollinators (Irwin 2000) it is important to consider flower color as populations of mixed color varieties suggest that some mechanism is allowing for the segregation of alleles along flower color. There is a difference between flower visitors and effective pollinators. It is possible that during pollinator surveys ineffective pollinators visiting plants are observed (Ollerton 2017). In the case of the two varieties of *T. erectum* it is possible that *Trillium* species are indeed attracting insects by their scent but also attracting other pollinators by their color. The ability of different traits to attract different pollinators has been reported in interactions between generalist plants and pollinator (Ollerton 2017). A previous study of pollen transfer between various *Trillium* taxa suggested that transfer of pollen was occurring between plants of different color but smaller numbers of pollen grains were transferred between different color individuals (Stoehrel 2010). This also suggests that pollinator selection acting on flower color is an incomplete barrier to gene flow (Stoehrel 2010), which may account for the presence of mixed

populations of various color flowers as well as populations of hybrid individuals exhibiting a blend of phenotypes.

Trillum erectum varieties are able to produce viable offspring by both outcrossing and self-pollinating as they have a week self- incompatibility system (Sage et al 2001.) This is also the case with cross pollination between both varieties regardless of flower color (Case and Case 1997). However seed sets are typically smaller in self-pollinated individuals (Irwin 2000). Thus cross pollination is essential to maintain a robust population (Sage et al. 2001).

Significance

This study will contribute to the general understanding of plant evolution focusing in the context of hybridization, more specifically how hybridizing taxa impact population dynamics in mixed and pure populations. This study also provides support for the ability of the use of simple methods of genetic analysis, the comparison of microsatellite regions, to generate meaningful data examining the subtle variations between individuals, populations, and genetically similar species.

Comparisons of allelic patterns in microsatellite loci in the taxa examined by this study should help to better understand the genetic relationships in the *T. erectum* species complex and to allow for better informed decisions to be made regarding the classification of the varieties of the complex. Knowledge of this nature is becoming more important as climate change continues to influence species range. At this time the species is not yet listed as threatened or endangered, but some varieties of the complex are marginalized and may continue to decline especially if introgression between larger populations of the species complex continues as their geographic rang shifts. In this study I present evidence that allelic data can be used to compare closely

related taxa in the *T. erectum* species complex. Furthermore, using population genetic analysis methods, I provide evidence that there is little genetic difference between *T. erectum* var. *erectum*, *T. erectum* var. *album*, and *T. rugelii*

CHAPTER TWO: METHODS

Tissue Collection

I collected 100 new tissue samples from four different locations in North Carolina and one location in Tennessee. This includes the sites: Harmon Den, Cold Spring Creek, Wolf Creek, Newfound Gap, and Fork Ridge Trail. Previously collected samples from four additional sites in North Carolina (WCU, Fleetwood, Dark Cove, and Balsam Mountain Preserve) were also included in the study (Table 1, Figure 1). An additional 20 tissue samples were added from a prior study at Western Carolina University, which were previously collected from Dark Cove Farm, a private property willing to participate in the study. Tissues were collected from a mixed taxa population. Samples were collected from both individuals that appear to represent a single taxon as well as individuals that exhibited traits of multiple taxa identified as hybrid forms. Voucher specimens were collected in areas where permission was obtained including the Wolf Creek, Harmon Den, and Cold Spring Creek populations. Vouchers were deposited in the Western Carolina University Herbarium (WCUH). Populations were found using herbarium records and then taxa were identified using a dichotomous key (Weakley 2020). The characteristics that distinguish *Trillium erectum* var. *erectum* from *Trillium erectum* var. *album* is simply the difference in color of flower petals (Weakley 2020). *Trillium rugelii* is distinguished from the two varieties of *Trillium erectum* by the presence of flowers that are held below the leaves both varieties of *Trillium erectum* have flowers above the level of the leaves (Weakley 2020). Hybrid individuals were classified by the presence of intermediate characteristics such as a blend of flower color or the presence of a bent peduncle rather than an erect or nodding peduncle.

Table 1. Locations of single and mixed taxa populations of *Trillium* species and varieties that were sampled for this study.

Single or Mixed Taxa	Taxa sampled	Site name	County, State	Abbreviation of sample set
Single Taxa	*T. erectum* var. *album*	Newfound Gap	Swain, NC	NFG_A
Single Taxa	*T. erectum* var. *erectum*	Fork Ridge Trail	Sevier, TN	FRT_E
Single Taxa	*T. erectum* var. *album*	Cold Spring Creek	Haywood, NC	CSC_A
Single Taxa	*T. erectum* var. *album*	Harmon Den	Haywood, NC	HD_A
Single Taxa	*T. rugellii*	Western Carolina University	Jackson, NC	WCU_R
Single Taxa	*T. erectum* var. *erectum*	Fleetwood	Ashe, NC	FW_E
Mixed Taxa	*T. erectum* var. *erectum*,	Wolf Creek	Jackson, NC	WLF_E
Mixed Taxa with hybrids	*T. erectum* var. *album*, *T.erectum* var. *erectum T. rugellii*	Dark Cove	Jackson, NC	DC_EAR
Mixed Taxa	*T. erectum* var. *album*, *T.erectum* var. *erectum T. rugellii*	Balsam Mountain Preserve	Jackson, NC	BMP_A BMP_E BMP_R

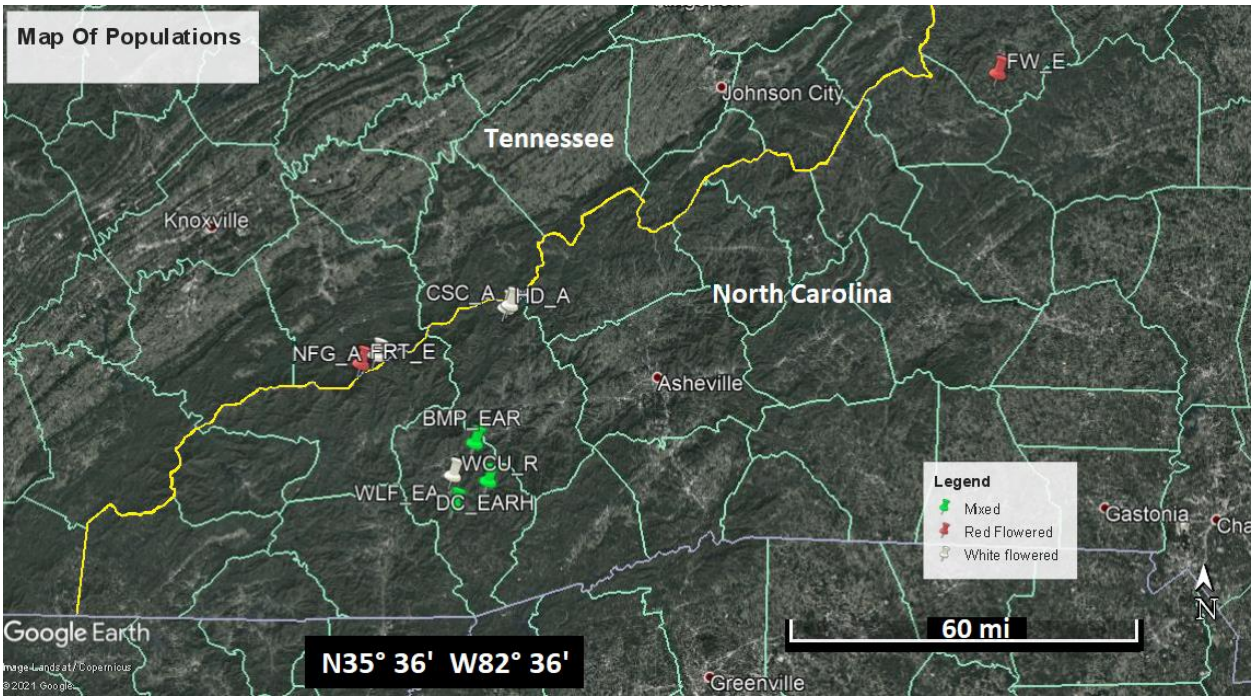


Figure 1. Map of western North Carolina marked with sample sites. Red pins indicate pure red flowered populations, white pins represent pure white flowered populations, and green pins represent populations mixed flower color varieties.

Leaf tissues were collected from 20 individuals of a single taxon at each location. 1cm by 1cm leaf punches were cut from individuals to cause minimal disturbance. Samples were stored on ice during collection and then stored at -80 °C until DNA was extracted.

DNA Extraction and Gene Amplification

Total DNA was extracted from frozen leaf tissue by grinding with a mortar and pestle in liquid nitrogen. An Omega Bio-Tek E.Z.N.A. Plant DNA Kit was used for the extraction of samples collected from the Wolf Creek, Harmon Den, Cold Spring Creek, Newfound Gap, and Fork Ridge Trail populations (Omega Bio-tek, inc. Norcross, Georgia). The sample set from Dark Cove Farm was collected prior to the other samples and was processed using a DNeasy Plant Mini kit (Qiagen Corporation, Valencia, California).

PCR (polymerase chain reaction) was performed using Omega Bio-tek 2x PCR Taq Mastermix (Omega Bio-tek, inc. Norcross, Georgia). Samples were prepared for 12.5 µl reactions which consisted of 6.25 µl of Omega Bio-tek 2x PCR Taq Mastermix, 5.25 µl H2O, 0.25 µl of forward primer, 0.25 µl of reverse primer, and 0.5 µl of undiluted template DNA. The microsatellite loci were targeted for amplification using TC69, TC48, and TC36 fluorescent dye-labeled primers developed for the sister taxon of the *Trillium erectum* species complex, *T. camchatcense* Ker Gawl (Kubota et al. 2006). Primer sequences used to amplify the TC36 loci were F: GTCCGAATAGTCGTCTGTCA, R:GCTTTGCATGGCAGGAACT, for the TC48 loci F:CAACCCGCAAGTATTTCAA, R:GAAATTAACTAAAGAAAGATTAGAGAGA, and for TC69 loci F:TTCATTACCCCTCGTCTCTC, R:CTCGTAGTGGAGTTGGAGAA (Kubota et al. 2006). Originally six primer pairs for six microsatellite loci were described, however this study uses only uses three loci as the additional three loci were not successfully amplified during preliminary testing of the samples.

PCR was performed using an Eppendorf Mastercycler Gradient (Eppendorf Hamburg, Germany) thermocycler. The cycling program used was a variant of the Touchdown PCR method, specifically Touchdown 60 (Korbie and Mattick 2008, Kubota et al. 2006). Conditions consisted of 20 cycles of denaturation for 30 seconds at 96 °C followed by a 30 second period of 60°C and a reduction in temperature from 60°C to -0.5 °C at a rate of 3 °C per second followed by a period of annealing at 72 °C for 30 seconds and an additional 30 cycles starting with 96 °C reducing to 48 °C and returning to 72 °C with a final period of 72 °C for 5 minutes.

Fragment Analysis

Electrophoresis was performed on an Applied Biosystems 3130xL capillary genetic analyzer (Applied Biosystems, Foster City, CA) to determine fragment length. This process was performed at the Western Carolina University Forensic Science Department. Prior to analysis, samples were mixed with Hi-Di Formamide and 500 ROX red size standard (Applied Biosystems) in a ratio of 8 µl Hi-Di Formamide, 0.5 µl Rox size, 0.75 µl PCR product template, and 0.75 µl H20. PCR product template was initially added in a quantity of 1.5 µl, however initial fragment analysis suggested that this concentration of template was excessive, and half of the quantity was replaced with H2O. After combining, the prepared samples were denatured at 95 °C for two minutes and immediately flash chilled on ice and placed into the genetic analyzer.

Fragment length analysis and genotype detection was performed using GeneMapper™ software version 4.0 (Applied Biosystems) to process electropherogram files generated during genetic analysis. Fragment length files were uploaded and analyzed using an allele-calling function in the GeneMapper program. This allele-calling function uses prespecified bins to call alleles that had peaks in the range of the bins. Bin sets were created in a previous project to identify alleles present in the samples. Additional bins were added based on the presence of allele peaks in multiple samples in the set. The allele-calling function generated diploid genotypes that were then manually inspected.

To this data set, an additional set of genotype data that had been collected in a prior Western Carolina University study was added. This new set consisted of fragment data from an additional 73 tissue samples collected from four unique geographic populations. These were processed in the same fashion as described previously in regard to collection, DNA extraction,

PCR, and fragment data processing, however PCR fragment electrophoresis was performed by GENEWIZ Corporation (Frederick, MD).

Statistical Analysis

Genotypes were downloaded as a .csv file from the GeneMapper program and loaded into an Excel file and formatted for GenAlEx (Peakall and Smouse 2006) analysis. The number of total alleles (N), number of effective alleles (Na), the expected heterozygosity (H_e), the Fixation index (F_{st}), Shannon's Information Index, Nei's genetic distance, were calculated using GenAlEx. Pairwise comparisons by population of Shannon's Information Index, Nei's genetic distance, and Fst values were calculated also using GenAlEx. These pairwise comparisons were then reloaded into GenAlEx and a Principle Coordinates Analysis (PCoA) was performed. A total genetic distance table was generated comparing each sample using the Genetic Distance calculation matrix in GenAlEx. This matrix was used for an Analysis of Molecular Variance (AMOVA). A Mantel test was performed to compare the pairwise comparisons of population genetic distance and population geographic distance using GenAlEx. GPS coordinates were first converted from UTM to distance by kilometer and a pairwise comparison table was generated using GenAlEx to compare distances between populations. The Mantel test compared the Nei's genetic distance to geographic distance for the 11 sample populations over 99 permutations.

Allelic data were formatted and entered into STRUCTURE (Pritchard et al. 2003), a Bayesian genetic clustering tool, and analyzed using the admixture model. The collection locations were used as priors in the analysis as recommended in the STRUCTURE User Manual for datasets with potentially weak signal, such as with a low number of genetic markers. Analysis was performed for assumed underlying genetic population sizes of 1 to 25. Parameters

of the analysis include a burn in period of 5000 repetitions before running 50,000 Markov Chain Monte Carlo (MCMC) repetitions for 20 separate iterations. The Structure Harvester Web v. 0.6.94 program (Earl & vonHoldt 2012) was used to determine the best value of genetic population size (K). The ad hoc examination of ΔK was used averaging the likelihood value of each assumed value of K (1-25) for each iteration (Earl and vonHoldt 2012).

Justifications of Methods

Based on Wright's F test, the Fst predicts that genetic differences among populations are related to the size of the populations and the amount of migration among populations. Larger populations among which it is common for individuals to migrate will have less genetic differentiation than smaller populations connected by less migration (Holsinger and Weir 2009).

The Fst statistic is a proportion between 0 and 1. An Fst of 0 represents all alleles the same. A low Fst value suggests that populations are breeding freely with large quantities of shared alleles between members of multiple populations. An Fst of 1 represents no overlap in allele frequency. A high Fst value suggest that populations are genetically isolated and share few alleles (Holsinger and Weir 2009). In a similar population genetic study of *Trillium govaniaum* Wall. Ex D. Don it was found that populations had an Fst value of 0.24 suggesting that there is a high amount of genetic variation between populations (Dhyani et al. 2020). Fst values can also be calculated for populations allowing for comparisons to be made between populations. It should be noted that Fst values will differ based on the variance of the loci being examined thus comparisons of different types of molecular markers the heterozygosity must be similar (Hedrick 1999). Also, if an allele is being selected for in its environment the Fst value will be larger in the selected allele among members of populations (Holsinger and Weir 2009). The statistic

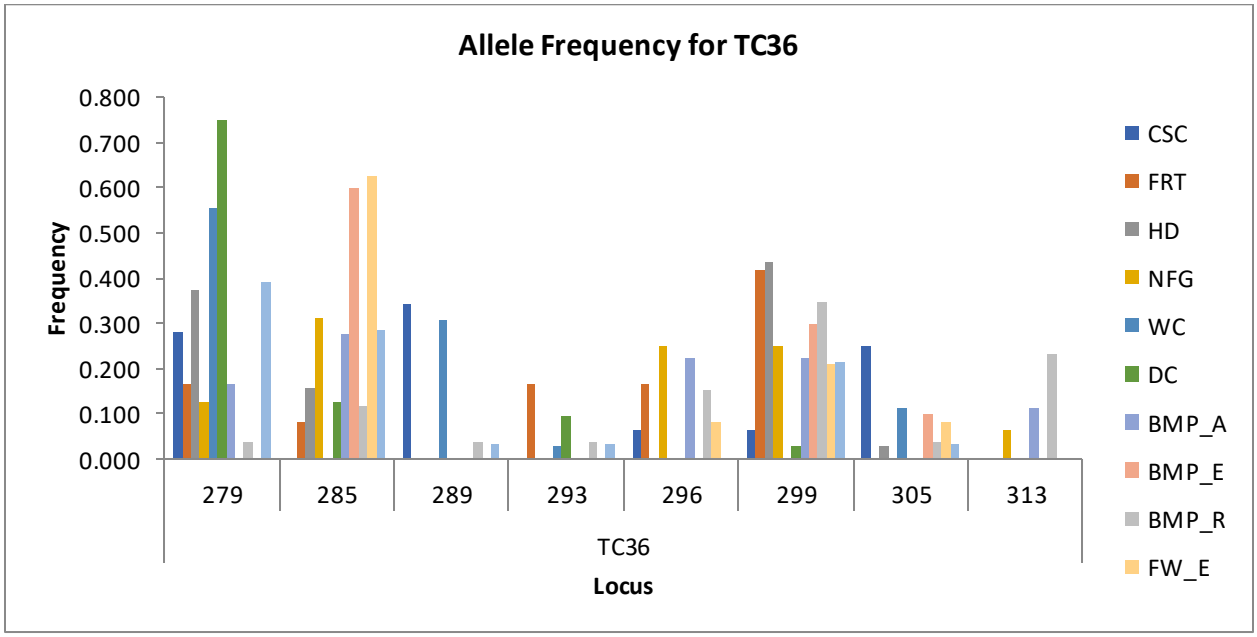
generated by this F test can be used to determine the goodness of fit with the Hardy-Weinberg equilibrium model (Holsinger and Weir 2009). The microsatellite loci are part of non-coding regions of DNA and are selected from various areas in the genome and therefore should not be under any selection pressure unless they are linked with genes under selection pressure.

The STRUCTURE 2.3.4 program (Pritchard et al. 2003) is a tool used to analyze genetic structuring in populations and examine the underlying genetic structure of populations by grouping individual samples based on shared patterns of genetic variation (Porras-Hurtado et al. 2013). The program uses a Bayesian statistical method utilizing Markov Chain Monte Carlo (MCMC) estimations to randomly assign individuals to groups and groups are then assessed and regrouped based on the amount of variation between individuals placed into the group. The method assumes a model in which there are K populations, each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are probabilistically assigned to one or more populations (more than one if their genotypes indicate that they are admixed) (Pritchard et al. 2003). Applying an admixture model, assuming that individuals may have mixed-population ancestry, is recommended for most natural populations because of its flexibility in finding structure if structure exists, rather than enforcing the stricter no-admixture model (Pritchard et al. 2003). In addition, prior population information can be applied, such as the LOCPRIOR, which uses sampling locations as prior information to assist the clustering. The LOCPRIOR model is recommended for use with data sets where the signal of structure is relatively weak, such as when few markers are used (Pritchard et al. 2003). Due to the small number of loci analyzed for each sample and the fact that I did know the location in which samples were collected, I used the location of the sample sites as a prior in my structure analysis. When running the analysis with no location prior added, no genetic structure was detected.

CHAPTER THREE: RESULTS

Allele Frequency

For the TC36 locus there were eight unique alleles that ranged from 279 to 313 base pairs, for the TC48 locus there were six unique alleles that ranged from 134-150 base pairs, and for the TC69 locus there were eleven unique alleles that ranged from 149-180 base pairs (Table 3). There are a few alleles that are very common across all populations such as alleles size 149 and 164 for the TC69 locus and 279, 285, and 299 for the TC36 locus (Figure 2).



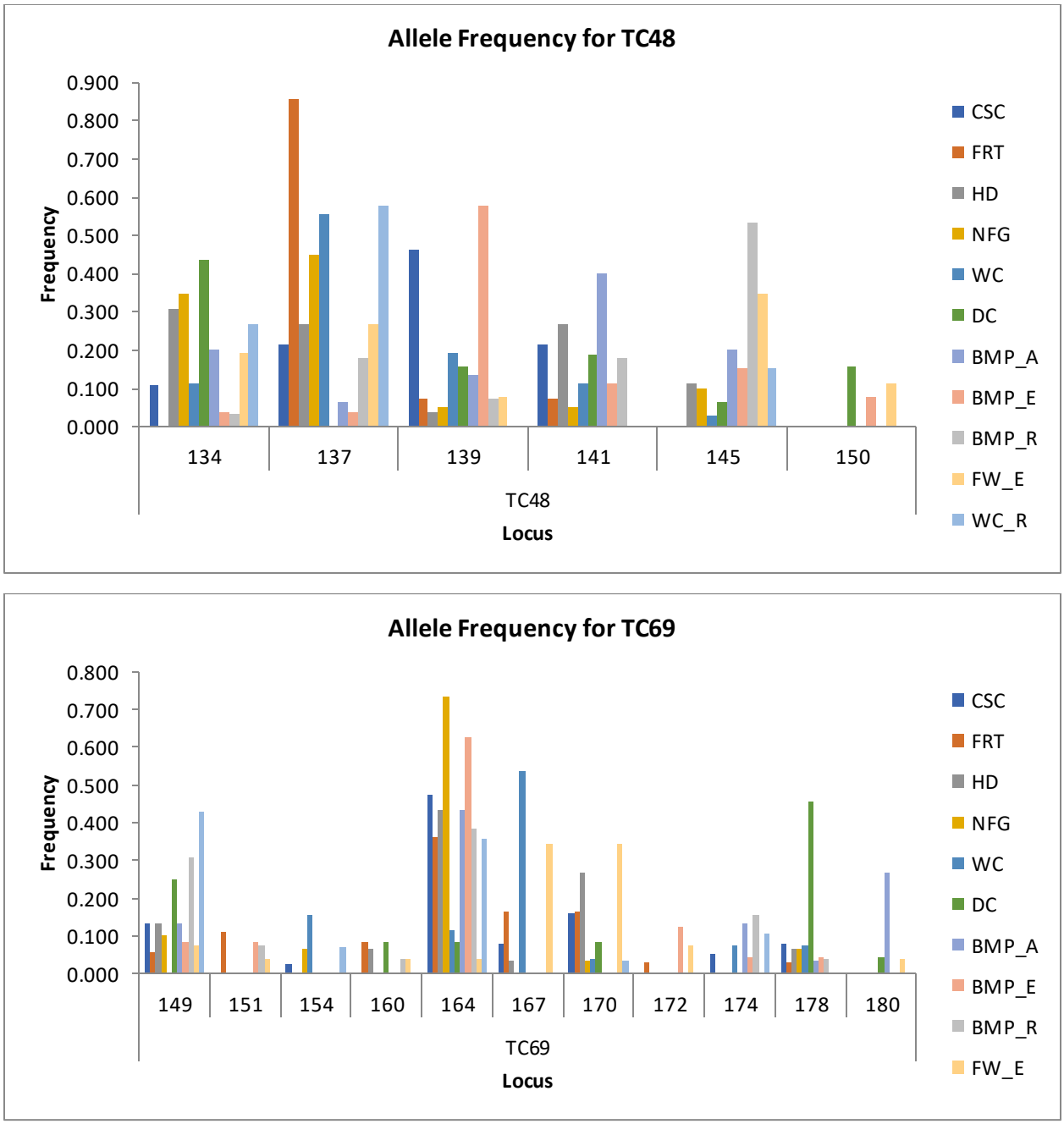


Figure 2. Frequency of alleles present in each population sample set (See table 1 for full names of sample sites) for each microsatellite locus. Numbers on the x-axis indicate allele size in nucleotide base-pairs.

Table 2. Microsatellite loci characteristics including repeat motif, primers used from Kubota et al. (2006), size range in base pairs, number of total alleles, observed heterozygosity (Ho) and expected heterozygosity (He) Hardy Weinberg Equilibrium P value (HWE).

Locus	Repeat motif	Primer sequence (5′-3′)	Size range (bp)	Number of Alleles	Ho	He	HWE (P value)
TC36	$(AG)_{11}$ $(AC)_8$ $(AG)_9$ $AC(AG)_8$	F:GTCCGAATAGTCGTCTGTCA R:GCTTTGCATGGCAGGAACT	279-313	8	0.296	0.658	0.000
TC48	$(TC)_{15}$ $TT(TC)_4$	F:CAACCCGCAAGTATTTCAA R:GAAATTAACTAAAGAAAGATTAGAGAGA	134-150	6	0.469	0.638	0.000
TC69	$(TC)_{10}$ $(AC)_6$ $(TC)_9$	F:TTCATTACCCCTCGTCTCTC R:CTCGTAGTGGAGTTGGAGAA	149-180	11	0.304	0.677	0.000

Genetic Structure Analysis

The Bayesian based STRUCTURE analysis showed the best fit number of underlying genetic population clusters to be represented by two clusters (k = 2). This was determined using the program Structure Harvester to analyze the change in likelihood method which generated the graph of the change in likelihood of k seen below (Figure 3) (Evanno et al. 2005, Dent and vonHoldt 2012). The amount of admixture varied between populations and displayed some correlation to geographic proximity but no directional trend (Figure 4). These two genetic clusters are best represented by the Wolf Creek population and The Balsam Mountain Preserve population although a degree of admixture is present in these populations with Wolf Creek displaying 1.41% admixture of the second genetic cluster and Balsam Mountain Preserve taxa

displaying admixture ranging from 6.5% to 8.05% (Table 4). All other populations showed a greater degree of admixture between the two genetic clusters. The population displaying the greatest amount of admixture between the two genetic structures is the Fork Ridge Trail population with 58.74% of cluster one and 41.26% of cluster two (Table 4).

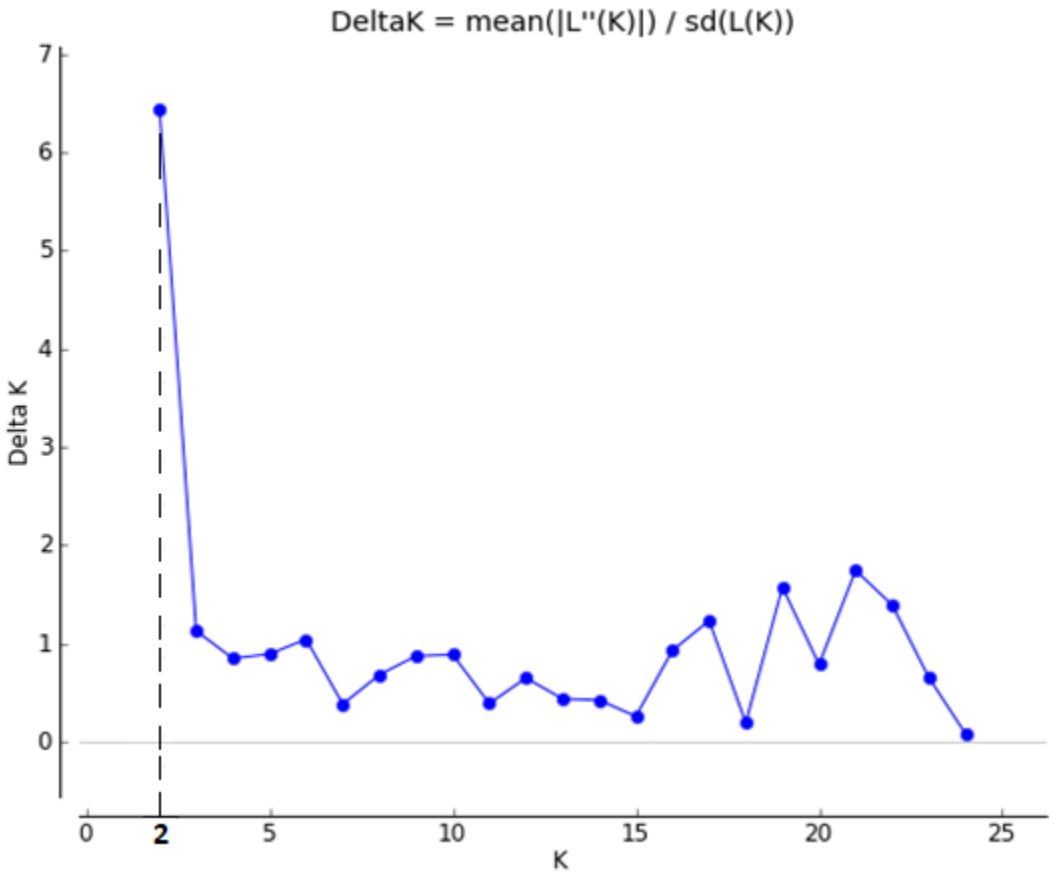


Figure 3. Delta K analysis graph of the output of the Evanno et al. (2005) method comparing changes in likelihood of k across different values of k sampled.

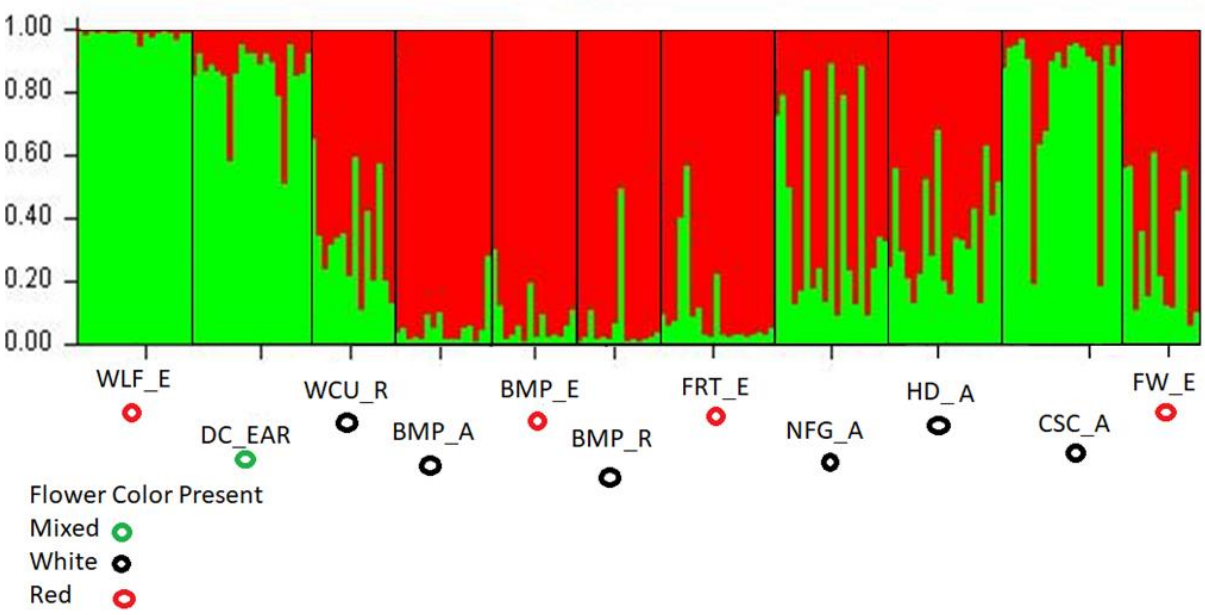


Figure 4. Bar graph from STRUCTURE analysis K=2 depicting percentage of admixture arranged by geographic location from most Southwest population to Northeast population.

Table 3. STRUCTURE percent admixture in populations

Sample Set	% Cluster 1	% Cluster 2
CSC_A	17.15	82.85
FRT_E	58.74	41.26
HD_E	65.1	34.91
NFG_E	89.51	10.49
WLF_R	1.41	98.59
DC_EAR	14.51	85.5
BMP_A	94.25	5.75
BMP_E	91.95	8.05
BMP_R	93.5	6.5
FW_E	69.4	30.6
WC_R	66.29	33.71

Principal Coordinates Analysis

Principal Coordinates Analysis of Nei's genetic distance indicated two groups with a few outliers that primarily grouped the white flowered taxa together and scattered the red and hybrid populations (Fig. 6). The Cold Spring Creek and Harmon Den were pure populations both consisting of white flowered individuals visually identified as *T. erectum* var. *album* and occurring nearby each other geographically. These populations were grouped with the Western Carolina University population, a pure population of white flowered individuals identified as *T. rugelii*. The other main group consisted of the white flowered individuals from the Balsam Mountain Preserve, *T. erectum* var. *album* and *T. rugelii,* and the Newfound Gap population of *T. erectum* var. *album*. However this group included the Fleetwood population which was a pure red *T. erectum* var. *erectum* population found furthest away from the other populations. Outlying population that do not appeared grouped include the Fork Ridge Trail population of *T. erectum* var. *erectum* that was closest geographically associated to the Newfound Gap population, the Wolf Creek population that consisted of *T. erectum* var. *erectum*, and the Dark Cove Farm population that consisted of all three taxa as well having visually identified hybrid variations. The red flowered *T. erectum* var. *erectum* from Balsam Mountain Preserve were also not as closely associated with the other two taxa sampled at the same geographic location but were placed closest to the other taxa sampled at the Balsam Mountain Preserve (Figure 6).

The PCoA analysis of the pairwise comparisons of Fst values is fairly similar to the PCoA of Nei's genetic distance with the exception of the placement of the Fleetwood population sample set which red flowered pure population of *T. erectum* var. *erectum* that is placed very

close to the white flowered pure populations of *T. erectum* var. *album* at the Newfound Gap site (Figure 7).

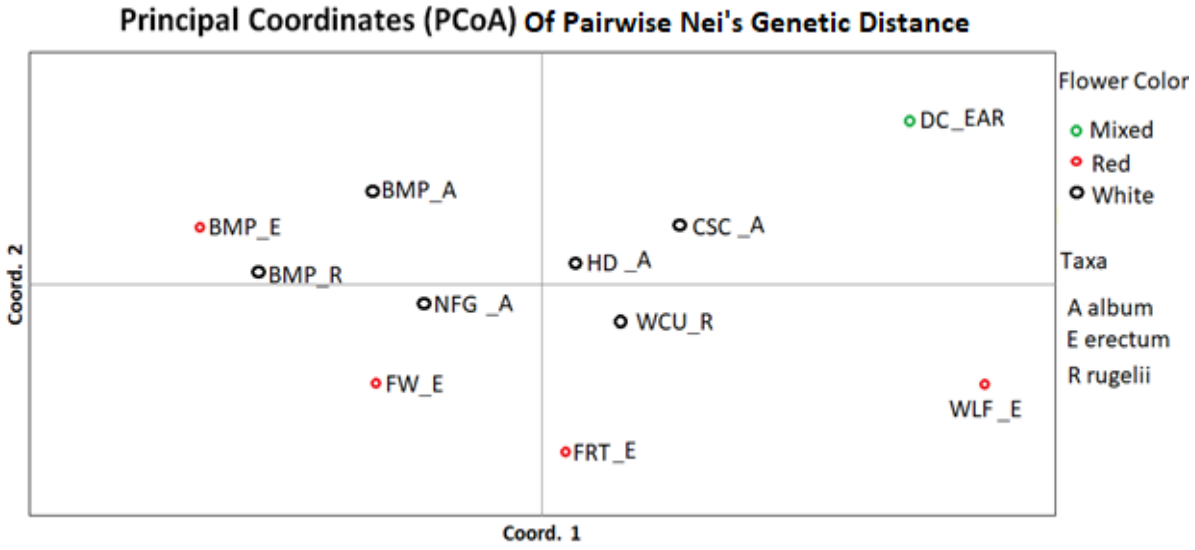


Figure 5. Principal Coordinates analysis of pairwise comparisons of Nei's Genetic Distance sample population

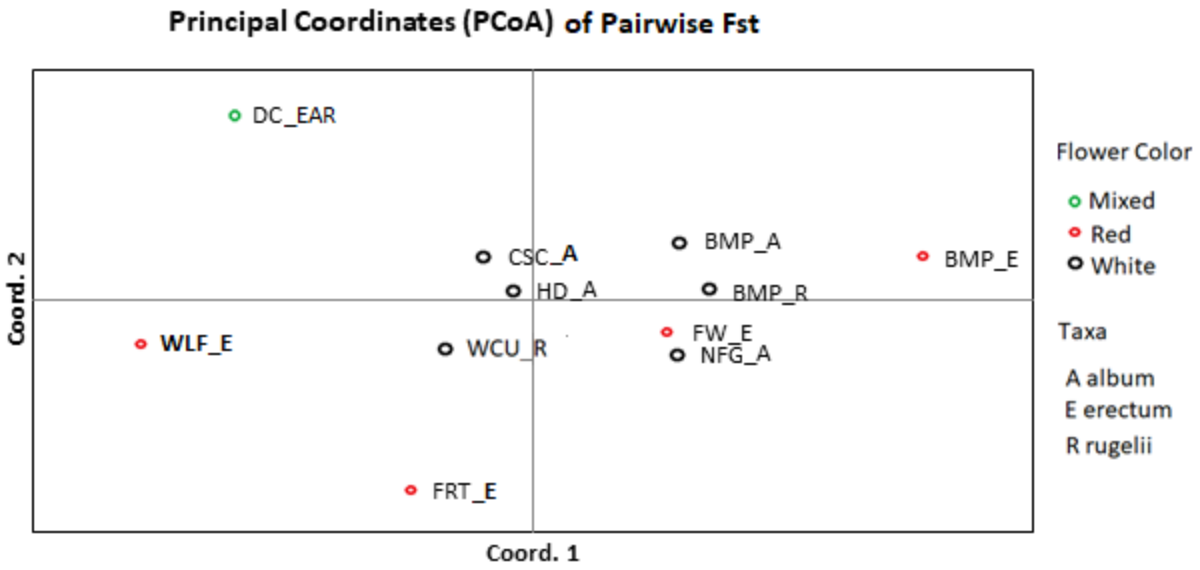


Figure 6. Principle Coordinates Analysis of the pairwise comparison of Fst values for sample populations.

AMOVA and Population Genetic Analyses

The AMOVA (analysis of molecular variance) suggests that 97% of the genetic variance is a result of variation within populations and that 3% of the total variance is a result of variation among populations (Table 5). The mean Fst value including all populations is 0.181 and Nm (number of migrants), calculated from the Fst value, is 1.144 and the inbreeding coefficient (Fis) is 0.455 (Table 6). Pairwise comparisons of Fst between populations suggest the greatest divergence between the Dark Cove Farm and the Fork Ridge Trail populations (Fst = 0.208) and least divergence between the *T. erectum* var. *album* and the *T. rugelii* taxa from the Balsam Mountain Preserve, (Fst = 0.044; Table 8). Pairwise Nei's genetic distance suggests the greatest divergence between the Wolf Creek population and the *T. erectum* var. *erectum* taxa at the Balsam Mountain Preserve (0.205) and the least divergence between the Newfound Gap and Western Carolina (0.799; Table 8).

Table 4. Analysis of molecular variance in all populations sampled.

Source	DF	SS	MS	Est. Var.	% variance
Among Pops	10	27.569	2.757	0.044	3%
Within Pops	363	455.610	1.255	1.255	97%
Total	373	483.179		1.299	100%

DF, Degree of freedom; SS, Sum of squares; MS, Mean square; Est. Var., Estimated variance.

Table 5. F-Statistics and Estimates of migration over All Pops for each Locus

All Pops.		TC69	TC48	TC36	Mean	SE
	Fis	0.551	0.264	0.550	0.455	0.096
	Fst	0.161	0.195	0.186	0.181	0.010
	Nm	1.306	1.035	1.091	1.144	0.083

Table 6. Population genetic parameters for each *Trillium* population sampled averaged across three microsatellite loci.

Population	N	Na	Ne	I	Ho	He	% P
CSC	16.333	5.333	3.471	1.412	0.319	0.710	100%
FRT	10.333	5.333	3.299	1.251	0.317	0.594	100%
HD	14.667	5.000	3.405	1.343	0.277	0.700	100%
NFG	13.667	5.000	2.954	1.226	0.253	0.620	100%
WC	16.333	5.000	2.679	1.219	0.336	0.624	100%
DC	14.667	5.000	2.890	1.231	0.340	0.613	100%
BMP_A	13.000	5.000	3.942	1.464	0.326	0.742	100%
BMP_E	10.000	5.000	2.392	1.144	0.302	0.579	100%
BMP_R	13.333	6.333	3.688	1.489	0.601	0.717	100%
FW_E	12.667	5.667	3.380	1.383	0.365	0.682	100%
WC_R	13.667	4.667	2.957	1.216	0.485	0.652	100%
Mean	13.515	5.212	3.187	1.307	0.357	0.658	100%

Na, Number of different alleles across all loci; Ne, Number of effective alleles; I, Shannon's information index; Ho, Observed heterozygosity; He, Expected heterozygosity; %P Percent polymorphism.

Table 7. Pairwise Population Fst values and Nei's Genetic Identity

CSC_A	FRT_E	HD_A	NFG_A	WLF_E	DC_EAR	BMP_A	BMP_E	BMP_R	FW_E	WC_R	
-	0.520	0.656	0.610	0.658	0.492	0.598	0.631	0.480	0.289	0.563	CSC_A
0.123	-	0.699	0.732	0.627	0.224	0.454	0.404	0.523	0.494	0.743	FRT_E
0.066	0.084	-	0.796	0.516	0.627	0.751	0.569	0.652	0.539	0.789	HD_A
0.086	0.086	0.052	-	0.421	0.379	0.751	0.677	0.666	0.520	0.799	NFG_A
0.078	0.101	0.110	0.146	-	0.497	0.311	0.205	0.255	0.394	0.600	WLF_E
0.118	0.208	0.093	0.163	0.128	-	0.483	0.270	0.283	0.263	0.552	DC_EAR
0.069	0.128	0.046	0.059	0.138	0.112	-	0.694	0.755	0.479	0.603	BMP_A
0.088	0.177	0.108	0.088	0.207	0.202	0.076	-	0.589	0.559	0.487	BMP_E
0.093	0.125	0.067	0.079	0.155	0.154	0.044	0.101	-	0.499	0.645	BMP_R
0.134	0.132	0.095	0.111	0.137	0.174	0.094	0.104	0.095	-	0.547	FW_E
0.093	0.077	0.048	0.055	0.096	0.113	0.081	0.135	0.076	0.099	-	WCU_R

Pairwise Fst values below diagonal and Nei's Genetic Identity values above diagonal.

Mantel Test

The Mantel tests comparing pairwise geographic distance to pairwise Fst values shows no significant correlation (Rxy = 0.11, p = 0.5) indicating that the analysis does not support the hypothesis that allele frequency among populations is correlated to geographic proximity (Figure 7). A Mantel test comparing the pairwise geographic distance to Nei's genetic distance also shows no significant correlation (Rxy = 0.166, p = 0.24) (Figure 8). Neither of these results suggest isolation by distance.

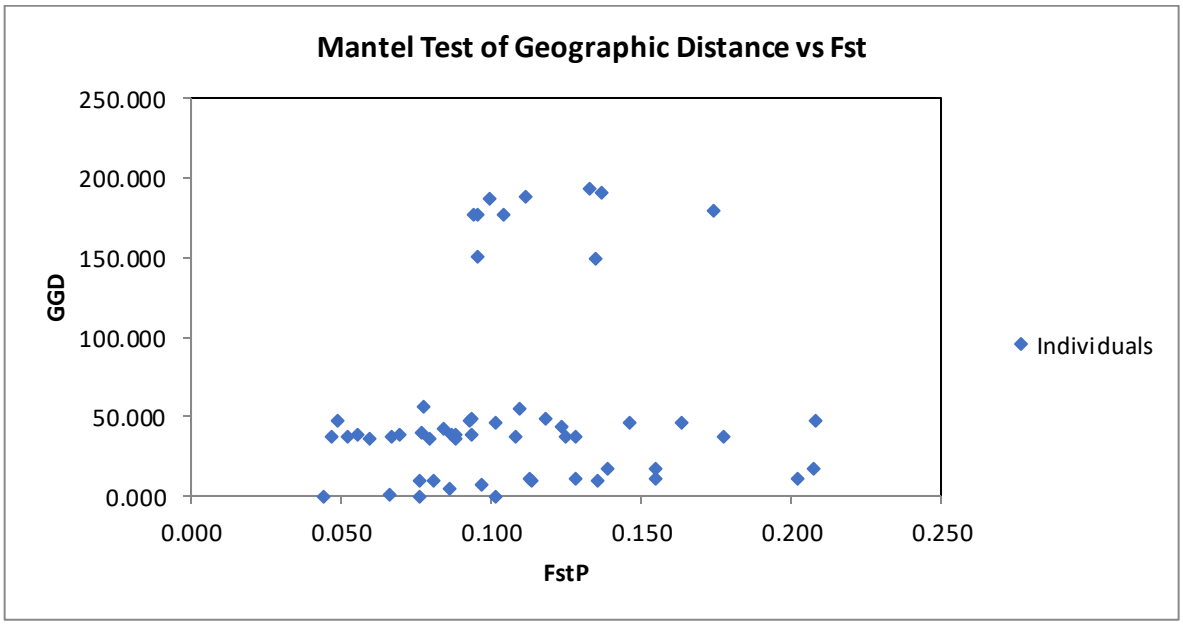


Figure 7. Graph of Mantel test comparing pairwise geographic distance in km to pairwise Fst.

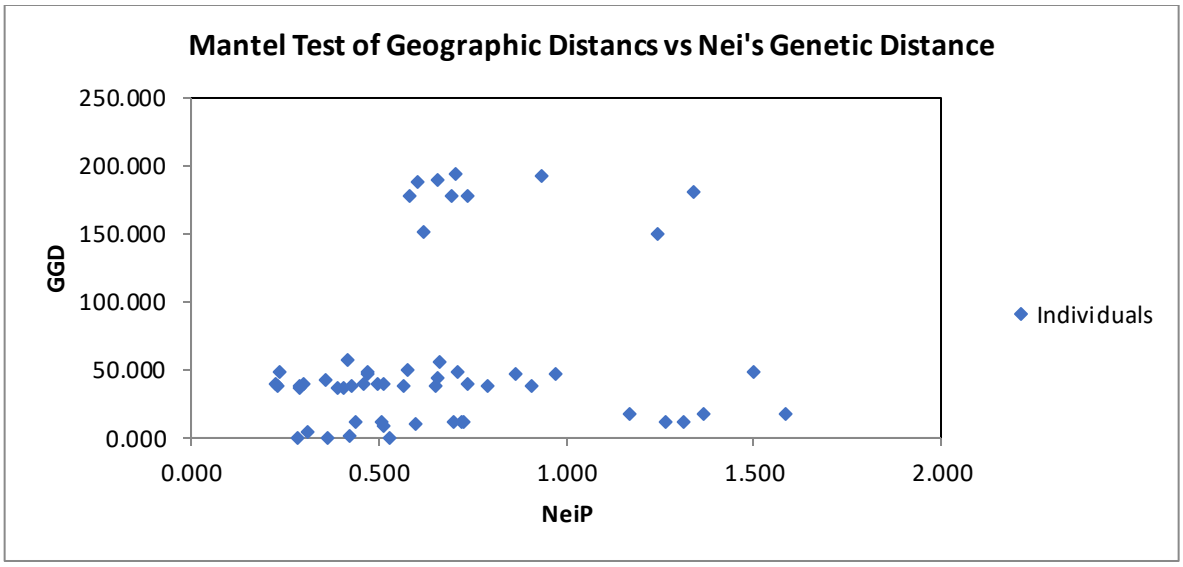


Figure 8. Graph of Mantel test comparing pairwise geographic distance in km to pairwise Nei's genetic distance.

CHAPTER FOUR: DISCUSSION AND CONCLUSION

Discussion

The objectives of this study were to determine what insight allelic data could provide on the population genetic relationships among the taxa specifically *T. erectum* var. *erectum*, *T. erectum* var. *album*, and *T. rugelii*, and second, to examine factors correlated with hybridization and gene flow such as geographic proximity, species identity, and is pollinator selection based on flower color and determine which is most influential.

Using microsatellites and calculating Fixation index values, I was able to calculate similarities among populations based on the proportion of shared alleles. Allelic data generated by the analysis of microsatellite loci is useful for detecting genetic similarity between closely related organisms without the use of more complex sequencing methods.

Pairwise comparisons of Fst values for populations can be used to assign a value that is descriptive of allele frequency in a population (Holsinger 2009). The pairwise comparisons of the Fst values for sites sampled suggest that the two most genetically similar taxa sampled were the *T. erectum* var. *album* and the *T. rugelii* collected at the Balsam Mountain Preserve area. The principle coordinates analysis based on the pairwise comparisons of Nei's genetic distance shows little or indistinct clustering, however populations and sample sets are arranged based on geographic distance (Figure 5). The sample sets of each taxa at Balsam Mountain Preserve form an implied cluster that is close to another implied cluster of the white flowered sampled taxa in pure populations. Red taxa sampled from pure populations are scattered around the white flowered group in a less discriminant fashion but the analysis does group the two pure red populations Fleetwood and Fork Ridge Trail regardless of the large distance, approximately 100

miles, between the two sites. This is evidence that there are similar numbers of alleles shared between pure populations of similar flower color regardless of geographic distance. This is consistent with the idea that these taxa are recently diverged (Stohrel 2010). The Dark Cove Farm population was an outlier placed far from the other clusters even though it contained a mix of the three taxa. This is contrary to the general assumption of the placement of populations of hybridizing individuals in a Principal Coordinates Analysis. Populations in which individuals are hybridizing share alleles and are expected to be clustered at a midpoint between the clusters of the other pure populations. This is not the case in my analysis as the Dark Cove Farm population is grouped away from any cluster. The isolation of this population and genetic drift and mutations at microsatellite loci could explain why the PCoA placed this population away from other pure populations. This evidence is corroborated by the genetic structure analysis.

The genetic structure analysis suggests that two underlying genetic clusters made up of an assortment of individuals with similar alleles are representative of the variation observed among the allele frequency of the populations. This is only when the location of the sample sites is included as a prior in the analysis as the location prior allows the use of populations to constitute the assortment of individuals. Without using location as a prior very little genetic structure is determined and a large best fit k value is reported.

The STRUCTURE analysis does indicate some degree of admixture in individuals of all populations, however, the Wolf Creek population is almost entirely indicative of one of the genetic clusters displaying very little admixture (Figure 4). The taxa sampled at the Balsam Mountain Preserve are indicative of the other underlying genetic cluster, as samples from this area display little admixture. The remaining populations show higher percentages of admixture up to 41% in the samples from the Fork Ridge Trail site (Table 4).There is no apparent reason

related to the differences in admixture of the populations i.e. distance, flower color, or number of taxa present in a population. This suggests that the genetic clusters may have been assigned based on the populations that showed the greatest differences in allele frequency among the populations. Populations with alleles in common with both of these outlying populations displayed the admixture. Evidence for this is in the pairwise comparison of Fst and Nei's genetic distance which indicated the populations with the greatest dissimilarity in alleles was the Wolf Creek populations and the Balsam Mountain Preserve population (Table 7).

In regard to my second question, what factor, geographic distance, flower color, or differences among taxa, is most influential of hybridization and gene flow, my study suggests geographic distance is most influential, and difference between taxa and flower color are not barriers to gene flow.

The pairwise comparisons of Fst and Nei's Genetic distance suggest that, as stated previously, *T. erectum* var. *album* and *T. rugelii* samples collected at the Balsam Mountain Preserve share the most alleles in common (Table 7). This means that two different taxa currently classified as separate species that share the same flower color and the same geographic locality are sharing more alleles than individuals of the same taxa found in different geographic localities. Also, the *T. erectum* var. *album* at the Balsam Mountain Preserve shared more genes with the *T. rugelii* than with the *T. erectum* var. *erectum* at BMP. This is evidence that both flower color and geographic proximity have influence on genes. However, the Mantel test does not show significant correlation between geographic distance and genetic distance. Taxa within close geographic proximity, i.e. the same population, are sharing more alleles than they share with other populations, but taxa sampled from various populations show no correlation between the sharing of alleles and the total distance between populations. Other direct evidence for this is

the observation of hybrid individuals found in mixed taxa populations. This suggests that individuals in close proximity regardless of flower color or taxon delimitation are able to produce fertile offspring.

Populations in my study show varying degrees of admixture that appear to be unrelated to the number of different taxa present. An example of this is the presence of very low percentages of admixture between the three taxa population at Balsam Mountain Preserve. Each of the sampled taxa show between 5.75 and 8.05% admixture which suggests that the difference in underlying genetic population is not a product of differences in flower color or taxa but rather geographic location. This is also the case in the Dark Cove Farm samples that show an average of 14.51% admixture even though three taxa are present and different color varieties are present.

The varying degrees of admixture may be a result of short dispersal of seeds and pollen (Table 3). Seeds are dispersed primarily by ants, often in short ranges, but have also been observed to be dispersed by deer. (Vellend et al 2003). However it is highly unlikely that and ant or deer would transfer a seed between populations that I have collected tissues from. Furthermore pollinators such as flies and beetles are also unlikely to carry pollen between populations that I have sampled.

Other studies of genetic structure between two hybridizing species show distinct genetic clusters for pure taxa with various rates of admixture in putative hybrids (Furches et al. 2013). This is not evident in my structure analysis as there are various amounts of admixture among populations with the hybrid population showing less admixture than pure populations. This indicates that the Dark Cove Farm population has its own unique genetic structure that is not made up of distinction between pure taxa and hybrid blends between the two taxa but rather a genetic structure representative of all taxa present sharing alleles in these populations. This

indicates that geographic proximity is more influential on genetic structure than flower color or taxa present as alleles are shared between individuals regardless of taxon or flower color.

A confounding aspect of the genetic analysis is the possibility that the microsatellite loci are linked to genes that are selected for such as genes for flower color. This is one of the reasons for selecting multiple microsatellite loci as the microsatellites were selected due to their diversity of location in the genome. Hardy Weinberg Equilibrium tests indicate that genetic variation will not remain constant regardless of outside factors such as selection. The p-values of the Chi^2 test indicate that none of the loci are in equilibrium in regard to the all populations (Table 2). This is likely due to inbreeding and may also be a result of the size of the populations.

The mean observed heterozygosity across all populations is low (Ho = 0.357) and the fixation index is low but each microsatellite locus was polymorphic at every population (Table 7). Though there are many individuals in the populations with homozygous allele pairs there is an abundance of variation among the individuals with these alleles. Though it is possible that this is due to allelic dropout in the microsatellite analysis as previous research shows that it is common for allele copies to fail to amplify (Wang et al. 2012). The low percentages of heterozygosity also suggests that there is a large amount of inbreeding resulting in the inbreeding coefficient (Fis = 0.455). This is consistent with previous pollination studies of *T. erectum* which indicate that, in general, pollinator service is unreliable because *T. erectum* flowers early in the spring when pollinators are less abundant and that self-incapability mechanisms in *T. erectum* are weak (Irwin 2002, Sage et al. 2001). Seeds are known to be dispersed by ants over relatively short distances, causing the nearest neighbors to be closely related (Case and Case 1997) High Fis values calculated from the comparison of heterozygosity to the fixation index suggests that inbreeding in populations sampled is occurring at a high rate. Furthermore *Trillium* reproduces

clonally by producing shoots from its rhizome (Case and Case 1997). When collection tissues I was careful to collect from plants that were not near each other, but cross pollination between clones would also decrease heterozygosity as no variation in genetic material would be inherited by the offspring.

Conclusion

The evidence presented in this study suggests that when analyzing three microsatellite loci present in each of the three taxa, from samples collected in the southern Appalachian Mountain region, there is little genetic difference among the taxa. This evidence agrees with prior studies using various genetic analysis techniques and provides a potential explanation for the disagreement between previous taxonomic classifications.

In my study high genetic diversity is indicated in the AMOVA where 97% of the variance is a result of variation within populations (Table 5). But why is there little evidence of genetic isolation between populations that are wide spread? It is highly unlikely that pollen or seeds are being dispersed among populations that are isolated by large geographic distance yet they share alleles. This suggest that the sharing of alleles is not a result of gene flow among populations but rather that genetic drift is acting slowly on these organisms or that not enough time has passed for an isolating mechanism to develop.

The Fixation index calculated shows no fixation of alleles between populations which suggests that either migration of alleles between populations is occurring or that the amount of divergence has not been enough for populations to show dissimilarity among the loci analyzed (Table 6). Given the geographic range of the populations selected, the poor rates of pollinator service reported in previous studies (Irwin 2000, Griffin and Barret 2004), and the low chance of

seed dispersal between populations due to the range of ant dispersal of seeds (Vellend et al. 2003), it is more likely that slow genetic drift is the reason for the presence of shared alleles between populations rather than gene flow among the populations. The lack of distinct trend in the pairwise comparison of genetic distance and the various amount of admixtures among populations are evidence of this. Each of these suggest that alleles are randomly mutating with little effect from genetic isolation. The genetic similarities of these taxa combined with their phenotypic variations makes it challenging to make a taxonomic classification that best represents the evolutionary linage.

The randomness of admixture in the genetic structure, the lack of defined clusters in the PCoA, and the lack of correlation between genetic distance and geographic distance suggest that genetic drift is the most likely cause of variation between populations. Based on these results it appears that these members of the *T. erectum* species complex is in the process of speciation and that secondary contact between different taxa is mitigating this effect.

A previous study, tracking pollen transfer between individuals using dyed pollen, suggested that pollen was being transferred between like colored flowers at a higher rate than between mixed colored flowers, although some pollen was being transferred between mixed colored individuals (Stoehrel 2010). This would explain the presence of populations with mixed floral colored individuals and hybrid mixed color individuals. The results of this study agree with this conclusion as we can see when comparing Fst and Nei's Genetic Distance populations with single flower color morphologies share more alleles in common than those that include mixed taxa (Table 8). However the Balsam Mountain Preserve samples, which represent multiple taxa in the same geographic area, indicate that alleles are shared among individuals regardless of taxa or flower color. This indicates that the differences between pure taxa populations of differing

flower color are most likely caused by geographic distance rather than the difference in flower color.

In a continuation of the study, it may be pertinent to analyze the *T. erectum* genome to determine exact sequences of microsatellite flanking regions to discover more microsatellite loci or better binding primers. This process has yielded informative results in two other studies of Asiatic *Trillium* species (Dhyani et al. 2020, Kubota et al. 2006). A study of this nature could provide useful evidence to resolve the taxonomy of the remaining taxa of the *T. erectum* species complex. In regard to the presence of mixed populations of varieties of *T. erectum*, a pollination observation study would provide insight on how the some populations maintain a strict barrier between color morphologies while others show evidence of cross pollination between the two. A study of this nature would aid in the understanding of how distinct color morphologies can remain in interbreeding populations.

In regard to the classification of *T. erectum* var. *erectum*, *T. erectum* var. album, and *T. rugelii* as species, the use of the Biological Species Concept would not recognize these taxa as species. Under this species concept all members in this complex should be considered to be a single species as each are known to be capable of interbreeding and overlap in geographic range and flowering phenology (Case and Case 1997, DeQueiroz 2007). Evidence presented in this study of genetic similarity among taxa suggest that according to the Biological Species Concept, *T. erectum* var. *erectum*, *T. erectum* var. *album*, and *T. rugelii*, should be classified as a single species. The Evolutionary Species Concept would also consider *T. erectum* var. *erectum*, *T. erectum* var. album, and *T. rugelii* as a single species because lineages of the species complex are not well defined due to the propensity of hybridization between members of the complex (Case and Case 1997, Wiley 1978). The Morphological Species Concept does differentiate

between the taxa of the *T. erectum* complex and previous classifications of species have been based on morphological characteristics (Barksdale 1939, Gleason 1906). However, the observation of hybrid forms with various morphological traits would confound the use of the Morphologic Species Concepts as each new hybrid form may be classified as a new species under this concept. Evidence presented in this study suggest that, *T. erectum* var. *erectum*, *T. erectum* var. album, and *T. rugelii*, may be better classified as subspecies of a single species.

LITERATURE CITED

Barksdale L. 1939. The pedicellate species of Trillium found in the Southern Appalachians. Journal of the Elisha Mitchell Society 54: 271-296.

Case F. W. *Trilliium* In: Flora of North America Editorial Committee, eds. 1993+. Flora of North America North of Mexico [Online]. 21+ vols. New York and Oxford. Vol. 3. http://beta.floranorthamerica.org/Fumaria. Accessed 5 mar. 2021.

Case F.W., Case R.B. 1997. *Trillium*. Portland, Ore: Timber Press.

Dequeiroz, K. 2007. Species Concepts and Species Delimitation Systematic Biology 56(6): 879-886.

Dhyani P., B. Sharma, P. Singh, M. Masand, R. Seth, R.K. Sharma. 2020. Genome-wide discovery of microsatellite markers and, population genetic dicersity inferences revealed high anthropogenic pressure on endemic population of *Trillium govanianum*. Industrial Crops & Products 154.

Earl, Dent A. and B.M. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4 (2): 359-361.

Evanno, G., S. Regnaut, J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology. 14 (8): 2611-2620.

Franz, N.M. and R.K. Peet. 2009. Perspectives towards a language for mapping relationships among taxonomic concepts. Systematics and Biodiversity 7(1): 5-20.

Furches M.S, R.L. Small, A. Furches. 2013. Hybridization leads to interspecific gene flow in *Sarracenia* (Sarraceniaceae). American Journal of Botany 100(10): 2085-2091.

Gleason, H. 1906. The Pedunculate Species of *Trillium*. Bulletin of the Torrey Botanical Club. 33: 387-396.

Griffin S. R., S.C.H. Barrett. 2004. Genetic variation in *Trillium erectum* (Melanthiaceae), a widespread forest herb in eastern North America. Canadian Journal of Botany 82: 316-321.

Hedrick P. W. 1999. Highly variable loci and their interpretation in evolution and conservation. Evolution 53(2): 313-318.

Holsinger, K.E., B.S. Weir. 2009. Genetics in geographically structured populations: defining, estimating and interpreting *F*st. Nature Reviews Genetics. 10(9): 639-650.

Ibrahim K.M., R.A. Nichols, G.M. Hewitt. 1995. Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. Heredity 77: 282-291.

Irwin, R. 2000. Morphological variation and female reproductive success in two sympatric *Trillium* species: evidence for phenotypic selection in *Trillium erectum* and *Trillium grandiflorum*. American Journal of Botany 87(2): 205-214.

Kato H., S. Kawano, R. Terauchi, M. Ohara, F.H. Uteh. 1995. Evolutionary Biology of *Trillium* and Related Genera (Trilliaceae) I. Restriction Site Mapping and Variation of Chloroplast DNA and its Systematics Implications. Plant Species Biology 10(1): 17-29.

Khosravi R., H.R. Rezaei, M. Kaboli. 2013. Detecting Hybridization between Iranian Wild Wolf (*Canis lupus* Pallipes) and Free-Ranging Domestic Dog (*Canis familiaris*) by Analysis of Microsatellite Markers. Zoological Science 30(1): 27-34 Zoological Society of Japan.

Korbie D.J., J.S. Mattick. 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nature Protocols 3: 1452-1456.

Kubota S., Y. Kameyama, M. Ohara. 2006. Characterization of six microsatellite markers in Trillium using a dual-suppression-polymerase chain reaction technique. Molecular Ecology Notes 6: 1135-1137.

McCormick A., Gray, T. and Mathews, K., 2018, March. Studying Hybridization within the Trillium erectum Complex via Microsatellite Amplification. In *79th Annual Meeting of ASB*. SEBIOLOGISTS.

Millam K.C. 2006. The Trillium Erectum Species-Complex (Melanthiaceae): Insights from Molecular Systematics and Biogeography, University of Wisconsin-Madison, Madison, Wisconsin.

Ollerton J. 2017. Pollinator Diversity: Distribution, Ecological Function, and Conservation. Annual Review of Ecology Evolutionary Systematics 48: 353-376.

Osaloo S.K., F.H. Utech, M. Ohara, S. Kawano. 1999. Molecular Systematics of Trilliaceae I. Phylogenetic Analyses of *Trillium* Using matK Gene Sequences. Journal of Plant Research 112: 35-49.

Page R.D.M., E.C. Holmes. 1998. *Molecular Evolution A Phylogenetic Approach*. Oxford: Blackwell Publishing.

Peakall R.O.D, Smouse, P.E. 2006. GENALEX6 genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes 6 (1): 288-295.

Porras-Hurtado L., Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M.V. Lareu. 2013. An overview of STRUCTURE: applications, parameter settings, and supporting software. Frontiers in Genetics. 4:98.

Pritchard J.K., W. Wen, D. Falush. 2003. Documentation for STRUCTURE Software: Version 2.3.

Sage T.L., S.R. Griffin, V. Pontieri, P. Drobac, W.W. Cole, S.C.H. Barrett. 2001. Stigmatic Self-Incompatibility and Mating Patterns in *Trillium grandiflorum* and *Trillium erectum* (Melanthiaceae) Annals of Botany 88(5): 829-841.

Stoehrel C.P. 2010. A Study of the Systematic Relationships between Members of the *Trillium erecutm* Complex. Western Carolina University, Cullowhee, North Carolina.

Todesco M., M.A. Pascual, G.L. Owens, K.L. Ostevik, B.T. Moyers, S. Hubner, S.M. Heredia, M.A. Hahn, C. Caseys, D.G. Bock, L.H. Rieseberg. 2016. Hybridization and extinction. Evolutionary Applications. 892-908.

Vellend M, Myers JA, Gardescu S, Marks PL. 2003. Dispersal of Trillium seeds by deer: Implications for long-distance migration of forest herbs. Ecology 84(4): 1067-1072.

Wang C, K. B. Schroeder, N. A. Rosenberg. 2012. A Maximum-Likelihood Method to Correct for Allelic Dropout in Microsatellite Data with No Replicate Genotypes. Genetics 192(2): 651-669.

Weakley A.S. 2020. Flora of the Southern United States, Edition of 20 October 2020. University of North Carolina Herbarium, North Carolina Botanical Garden, Chapel Hill, NC.

Wiley E. O. 1978. The Evolutionary Species Concept Reconsidered. Systematic Biology 27: 17-26.

Wu Y., T. Zhong, Z. Zhang, Q. Li. 2018. Pollinator-mediated selection on floral traits varies in space and between morphs in *Primula secundiflora*. Annals of Botany 10: 59.

Yan L., K.S. Burgess, R. Milne, C. Fu, D. Li, L. Gao. 2017. Asymmetrical natural hybridization varies among hybrid swarms between two diploid *Rhododendron* species. Annals of Botany 120: 51-81.

Zaya D.N., S.A. Leicht-Young, N.B. Pavlovic, K. A. Feldheim, M.V. Ashley. 2015. Genetic characterization of hybridization between native and invasive bittersweet vines (*Celastrus* spp.). Bio Innovation 17: 2975-2988.

APPENDICES

Apendix A: Trillium Key Used (Weakley 2020).

Trillium erectum group

Petals relatively thick in texture, straight-margined, maroon or white, rarely yellow or green (if white, turning brown with age); stigmas thicker at base, tapering gradually toward tip, distinct; ovary purple-black, maroon, pink, or white, 6-angled; [*Erectum* group]

Trillium erectum var. *erectum*

Flowers held at or above the level of the leaves (the pedicel nearly horizontal, inclined above the horizontal, or erect).

> Ovary globose, widest near the middle, black to purplish black; petals white, maroon, yellowish, or otherwise.
>
> > Petals lanceolate to narrowly ovate or elliptic, spreading from base in the same plane as the sepal, rarely > 2 X as broad as the sepals; sepals 0.5-0.8 X as long as the pedicel, weakly sulcate-tipped roundly keeled and upturned near apex); flower fragrance unpleasant, musty.
> >
> > > Petals maroon to red

Trillium erectum var. *album*

Flowers held at or above the level of the leaves (the pedicel nearly horizontal, inclined above the horizontal, or erect).

> Ovary globose, widest near the middle, black to purplish black; petals white, maroon, yellowish, or otherwise.
>
> > Petals lanceolate to narrowly ovate or elliptic, spreading from base in the same plane as the sepal, rarely > 2 X as broad as the sepals; sepals 0.5-0.8 X as long as the pedicel, weakly sulcate-tipped roundly keeled and upturned near apex); flower fragrance unpleasant, musty.
> >
> > > Petals white

Trillium rugelii

Flowers held below the leaves (the pedicel declined below a horizontal plane).

> Stamens at most 1.5 X longer than the pistil, filaments shorter than the ovary, white (less commonly purplish), the anther sacs lavender to vivid purple (or albino); ovary white to pink or dull red, large, ovoid, 10-17 mm long; flower fragrance various; pedicel short to long, 1.5-12 cm long; petals not strongly overlapping, usually white (rarely maroon).
>
> > Pedicels short, 1.5-4 cm long.
> >
> > > Anthers 7.0 mm long or more, longer than the filaments; petals ovate to elliptic, much broader than the sepals; [of mesic forests of n. NC. southward]

Appendix B: Genotype Table.

Genotype table for all samples.

Sample	Pop	TC69		TC48		TC36	
CSC_1_69	CSC	164	164	141	141	279	279
CSC_10_69	CSC	149	170	137	137	289	305
CSC_11_69	CSC	149	167	139	139	289	289
CSC_12_69	CSC	167	178	137	139	305	305
CSC_13_69	CSC	164	164	141	141	289	305
CSC_15_69	CSC	164	164	139	139	296	296
CSC_16_69	CSC	164	164	0	0	0	0
CSC_17_69	CSC	164	164	139	139	0	0
CSC_18_69	CSC	174	178	134	139	0	0
CSC_19_69	CSC	170	174	0	0	279	279
CSC_20_69	CSC	149	178	139	139	0	0
CSC_2_69	CSC	170	170	0	0	289	305
CSC_3_69	CSC	170	170	134	139	289	289
CSC_14_69	CSC	154	164	139	139	279	279
CSC_4_69	CSC	164	164	0	0	289	289
CSC_5_69	CSC	164	164	137	137	305	305
CSC_6_69	CSC	149	149	0	0	299	299
CSC_7_69	CSC	0	0	134	137	289	289
CSC_8_69	CSC	164	164	0	0	279	305
CSC_9_69	CSC	164	167	141	141	279	279
FRT_1_69	FRT	164	164	137	137	279	279
FRT_10_69	FRT	167	170	0	0	0	0

FRT_11_69	FRT	170	170	0	0	0	0
FRT_12_69	FRT	170	170	0	0	296	296
FRT_13_69	FRT	151	151	0	0	0	0
FRT_14_69	FRT	167	167	0	0	0	0
FRT_15_69	FRT	151	164	0	0	0	0
FRT_16_69	FRT	149	164	0	0	0	0
FRT_17_69	FRT	164	172	0	0	0	0
FRT_18_69	FRT	167	167	137	137	0	0
FRT_2_69	FRT	164	164	137	137	285	299
FRT_3_69	FRT	0	0	0	0	293	293
FRT_19_69	FRT	151	164	137	137	0	0
FRT_4_69	FRT	160	160	0	0	299	299
FRT_5_69	FRT	167	178	137	141	0	0
FRT_6_69	FRT	164	170	0	0	299	299
FRT_7_69	FRT	149	164	0	0	0	0
FRT_8_69	FRT	160	164	137	137	0	0
FRT_9_69	FRT	164	164	137	139	0	0
HD_1_69	HD	0	0	137	137	285	285
HD_10_69	HD	164	164	134	139	279	279
HD_11_69	HD	170	170	137	145	279	299
HD_12_69	HD	170	170	145	145	279	299
HD_13_69	HD	164	164	0	0	299	299
HD_14_69	HD	170	170	137	137	299	299
HD_15_69	HD	164	164	0	0	279	279
HD_16_69	HD	149	167	0	0	299	299
HD_17_69	HD	178	178	141	141	0	0
HD_18_69	HD	170	170	134	141	299	299
HD_19_69	HD	149	149	137	137	299	299
HD_20_69	HD	160	160	134	141	0	0
HD_3_69	HD	149	164	141	141	285	305
HD_4_69	HD	164	164	134	134	0	0
HD_5_69	HD	0	0	134	134	279	285
HD_6_69	HD	164	164	0	0	299	299
HD_7_69	HD	0	0	0	0	279	279
HD_8_69	HD	0	0	134	141	279	285
HD_2_69	HD	164	164	0	0	279	279
NFG_1_69	NFG	164	164	134	134	279	296
NFG_10_69	NFG	164	170	134	137	285	285
NFG_11_69	NFG	149	149	0	0	0	0
NFG_12_69	NFG	164	164	137	137	279	279
NFG_13_69	NFG	178	178	0	0	0	0

NFG_14_69	NFG	164	164	0	0	279	296
NFG_15_69	NFG	164	164	137	137	0	0
NFG_16_69	NFG	164	164	137	139	299	299
NFG_17_69	NFG	164	164	0	0	313	313
NFG_19_69	NFG	154	154	137	137	299	299
NFG_2_69	NFG	164	164	0	0	285	285
NFG_20_69	NFG	149	164	0	0	296	296
NFG_3_69	NFG	164	164	0	0	285	285
NFG_4_69	NFG	164	164	134	134	296	296
NFG_5_69	NFG	164	164	134	145	296	296
NFG_6_69	NFG	0	0	0	0	299	299
NFG_7_69	NFG	0	0	0	0	299	299
NFG_8_69	NFG	0	0	137	145	285	285
NFG_9_69	NFG	0	0	134	141	285	285
WC_1_69	WC	167	167	137	139	289	289
WC_10_69	WC	0	0	0	0	279	279
WC_11_69	WC	167	167	137	139	279	279
WC_17_69	WC	0	0	137	139	289	289
WC_12_69	WC	167	167	137	137	279	279
WC_13_69	WC	167	167	137	139	279	279
WC_18_69	WC	0	0	137	139	279	279
WC_14_69	WC	167	167	137	137	305	305
WC_15_69	WC	167	167	137	137	279	279
WC_16_69	WC	167	167	137	139	279	279
WC_2_69	WC	154	154	137	145	305	305
WC_3_69	WC	154	164	137	141	289	289
WC_4_69	WC	164	164	134	141	289	293
WC_5_69	WC	154	174	134	141	289	289
WC_6_69	WC	178	178	134	137	289	289
WC_7_69	WC	0	0	137	137	279	279
WC_8_69	WC	170	174	137	139	0	0
WC_9_69	WC	0	0	137	137	279	279
WC_20_	WC	0	0	134	141	279	279
DC_4	DC	0	0	134	141	0	0
DC_20	DC	0	0	134	141	279	279
DC_14	DC	149	149	141	150	279	279
DC_6	DC	178	178	0	0	279	285
DC_3	DC	170	170	141	150	0	0
DC_2	DC	0	0	134	141	0	0
DC_17	DC	149	149	0	0	279	299
DC_13	DC	0	0	150	150	0	0

DC_9	DC	178	178	139	139	279	293
DC_1	DC	0	0	139	150	279	279
DC_10	DC	0	0	134	139	279	293
DC_11	DC	164	164	134	134	279	279
DC_18	DC	0	0	134	134	279	279
DC_19	DC	178	178	134	134	279	285
DC_12	DC	178	178	134	139	285	285
DC_7	DC	160	160	145	145	279	279
DC_5	DC	178	178	0	0	279	279
DC_15	DC	178	180	134	134	279	293
DC_8	DC	149	149	0	0	279	279
DC_16	DC	0	0	134	141	279	279
BMPA10_36-69.fsa	BMP_A	180	180	141	141	0	0
BMPA11_36-69.fsa	BMP_A	149	164	141	141	0	0
BMPA12_36-69.fsa	BMP_A	180	180	141	141	285	285
BMPA13_36-69.fsa	BMP_A	180	180	139	141	285	285
BMPA14_36-69.fsa	BMP_A	180	180	137	145	313	313
BMPA15_36-69.fsa	BMP_A	149	174	145	145	279	279
BMPA16_36-69.fsa	BMP_A	149	174	141	145	279	285
BMPA1_36-69.fsa	BMP_A	0	0	139	139	0	0
BMPA2_36-69.fsa	BMP_A	164	164	134	141	296	296
BMPA3_36-69.fsa	BMP_A	164	164	134	141	296	296
BMPA4_36-69.fsa	BMP_A	164	164	134	134	299	299
BMPA5_36-69.fsa	BMP_A	164	164	134	139	0	0
BMPA6_36-69.fsa	BMP_A	149	164	134	137	0	0
BMPA7_36-69.fsa	BMP_A	164	174	145	145	299	299
BMPA8_36-69.fsa	BMP_A	164	164	0	0	0	0
BMPA9_36-69.fsa	BMP_A	174	178	141	141	0	0

BMPE12-36_69.fsa	BMP_E	174	178	139	139	0	0
BMPE13-36_69.fsa	BMP_E	151	151	139	150	0	0
BMPE16-36_69.fsa	BMP_E	172	172	139	145	0	0
BMPE17-36_69.fsa	BMP_E	164	172	0	0	0	0
BMPE19-36_69.fsa	BMP_E	149	149	139	139	0	0
BMPE2-36_69.fsa	BMP_E	164	164	145	145	299	299
BMPE20-36_69.fsa	BMP_E	0	0	137	150	0	0
BMPE3-36_69.fsa	BMP_E	164	164	139	141	285	285
BMPE4-36_69.fsa	BMP_E	164	164	139	141	299	305
BMPE5-36_69.fsa	BMP_E	164	164	139	139	285	285
BMPE6-36_69.fsa	BMP_E	164	164	139	145	0	0
BMPE7-36_69.fsa	BMP_E	164	164	134	141	285	285
BMPE8-36_69.fsa	BMP_E	164	164	139	139	0	0
BMPE9-36_69.fsa	BMP_E	0	0	139	139	0	0
BMPR10-36_69.fsa	BMP_R	149	164	145	145	285	299
BMPR13-36_69.fsa	BMP_R	149	164	141	145	0	0
BMPR14-36_69.fsa	BMP_R	151	151	137	145	289	296
BMPR17-36_69.fsa	BMP_R	164	164	141	145	285	313
BMPR18-36_69.fsa	BMP_R	164	164	134	139	299	299
BMPR2-36_69.fsa	BMP_R	149	149	137	145	299	313
BMPR20-36_69.fsa	BMP_R	160	178	145	145	285	313
BMPR21-36_69.fsa	BMP_R	0	0	141	141	293	305
BMPR3-36_69.fsa	BMP_R	149	174	145	145	299	313

BMPR4-36_69.fsa	BMP_R	174	174	137	145	299	313
BMPR6-36_69.fsa	BMP_R	164	174	145	145	299	299
BMPR7-36_69.fsa	BMP_R	149	149	139	145	296	296
BMPR8-36_69.fsa	BMP_R	149	164	137	137	296	313
BMPR9-36_69.fsa	BMP_R	164	164	141	145	279	299
FW1.fsa	FW_E	167	167	139	150	285	285
FW2.fsa	FW_E	167	167	137	150	285	285
FW3.fsa	FW_E	170	170	145	145	285	285
FW4.fsa	FW_E	167	170	134	134	285	299
FWE10.fsa	FW_E	149	149	139	145	0	0
FWE11.fsa	FW_E	160	170	137	137	305	305
FWE12.fsa	FW_E	170	170	134	150	296	299
FWE13.fsa	FW_E	151	170	137	145	285	285
FWE5-36_69.fsa	FW_E	164	172	134	134	285	285
FWE6-36_69.fsa	FW_E	167	167	137	145	285	285
FWE7-36_69.fsa	FW_E	167	167	137	137	285	285
FWE8.fsa	FW_E	172	180	145	145	296	299
FWE9.fsa	FW_E	170	170	145	145	299	299
WC3.fsa	WC_R	164	170	134	134	279	293
WC4.fsa	WC_R	149	149	137	137	279	285
WC7.fsa	WC_R	149	154	137	137	299	299
WC8.fsa	WC_R	149	149	134	134	279	285
WCR1-36_69.fsa	WC_R	149	164	137	137	279	285
WCR12.fsa	WC_R	149	149	134	137	289	299
WCR13.fsa	WC_R	149	149	137	145	279	285
WCR14.fsa	WC_R	164	174	137	137	279	279
WCR15.fsa	WC_R	174	174	137	145	299	299
WCR2-36_69.fsa	WC_R	149	154	134	134	299	305
WCR5-36_69.fsa	WC_R	164	164	137	145	279	285
WCR6-36_69.fsa	WC_R	149	164	137	137	279	279
WCR_10.fsa	WC_R	164	164	137	145	279	285
WCR_9.fsa	WC_R	164	164	0	0	285	285